



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/944,332	08/30/2001	Liqin Shen	JP920000191US1 (590.079)	1865
35195	7590	02/09/2006	EXAMINER	
FERENCE & ASSOCIATES 409 BROAD STREET PITTSBURGH, PA 15143			HAN, QI	
			ART UNIT	PAPER NUMBER
			2654	

DATE MAILED: 02/09/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<p align="center"><b>Office Action Summary</b></p>	<b>Application No.</b> 09/944,332	<b>Applicant(s)</b> SHEN ET AL.	
	<b>Examiner</b> Qi Han	<b>Art Unit</b> 2654	

**-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --**

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 21 September 2005.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-19 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-19 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☒ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |  |  |
|--|--|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)<br>2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)<br>3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____<br>5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)<br>6) <input type="checkbox"/> Other: _____ |
|--|--|

### **DETAILED ACTION**

1. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

#### ***Response to Amendment***

2. This communication is responsive to the applicant's amendment dated 9/21/2005.

The examiner withdraws the disclosure objection b, because the applicant amended Table 1 in the specification.

The examiner withdraws the rejection of claims 1 and 10 regarding claimed limitation of “a cleaned corpus”, under 35 U.S.C. 112, 2<sup>nd</sup>, because the applicant further clarifies/explains the limitation (see the amendment: page 9), which can be interpreted in a broad sense (see the claim rejection under 35 U.S.C. 103 below).

#### ***Response to Arguments***

3. Applicant's arguments filed on under 35 U.S.C. 112 with respect to claims 1-19 have been fully considered but they are not persuasive.

In response to applicant's arguments with respect to claim 1 (also related to claims 10 and 19) that “neither Wang (primary reference) nor Razin (secondary reference), nor the combination of the two, teach or suggest the limitations of the instant invention”, “not only is there no motivation to combine the references, no expectation of success, but actually combining the

Art Unit: 2654

references would not produce the claimed invention” because Razin “uses a tree based on stemmed words and known elements, not character strings” (the amendment: page 11, paragraph 1 to page 12, paragraph 1), the examiner respectfully disagrees with the applicant’s arguments and has a different view of prior art teachings and claim interpretations.

Firstly, by reviewing the claim rejection and the cited references, the examiner believes that the previous prior art rejection is proper, because the combined references cover all claimed limitations and obviousness/motivation of combining the cited references (see detail in the claim rejection).

Secondly, in response to applicant's argument that “combining Wang and Razin would result in producing a language model of phrases which includes a lexicon of standard phrases rather than words...” (the amendment: page 12, paragraph 1), the test for obviousness is not whether the features of a secondary reference may be bodily incorporated into the structure of the primary reference; nor is it that the claimed invention must be expressly suggested in any one or all of the references. Rather, the test is what the combined teachings of the references would have suggested to those of ordinary skill in the art. See *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981). In this case, it would have been obvious to one of ordinary skill in the art at the time the invention was made to recognize that filtering a list of words is similar to filtering a list of phrases, because a phrase, with plain meaning, “**a word** or group of words forming a syntactic constituent with a signal grammatical function” (Merriam-Webster’s Collegiate dictionary, 10<sup>th</sup> edition, page 875), which means that a list of words can be broadly interpreted as a list of phrases. Further, it is noted that the applicant himself, in fact, defines a “new word” as

“a consecutive character string which occurs at least K times in given corpus” (specification: page 11, lines 8-9), which can be read on a phrase.

Furthermore, in response to applicant's arguments against the references individually that “Wang does not expressly disclose filtering... **nor does Wang simply output words**”(the amendment: page 10, paragraph 3) and “Razin outputs standard phrases, not new words in a document” (the amendment: page 11, paragraph 1), one cannot show nonobviousness by attacking references individually where the rejections are based on combinations of references. See *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981); *In re Merck & Co.*, 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986). It is noted that, Wang discloses, as stated in the claim rejection, ‘a textual corpus is dissected (interpreted as split) into a plurality of items (sub strings)’ and ‘counts the number of occurrences of a particular item (word, character, etc.)’, and ‘counting the occurrence of strings of characters of a sequence of words’ (corresponding new words) (column 1, lines 35-62), which suggests that the Wang’s discourse has capability of outputting the strings of the characters (or new words). It is also pointed out that, the reason to introduce secondary reference (Razin) is to combine Wang with the **expressly** disclosed feature of **filtering** a list from Razin. Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to recognize the combined system from the prior art disclosure has capability of implementing the claimed functionality and structure as claimed (see detail in the claim rejection below). In addition, the examiner slightly modifies the claim rejection for the purpose of reflecting the applicant’s arguments on this issue, without changing ground.

The response to applicant's arguments regarding claims 4-5 and 13-14 (amendment: page 12, paragraph 2 to page 13, paragraph 1) is directed to the response for claim 1 described above

Art Unit: 2654

and the corresponding claim rejection below, because applicant's arguments either are based on the same issue as claim 1 or simply request for reconsideration without further addressing specific issue.

### *Specification*

4. The disclosure is objected to because of the following (use the same reference numbers):
  - a. On page 6, line 7, regarding the content “length (S)-N is  $N(N+1)/2$ ”, even though the applicant amended it as “length (S)=N is  $N(N+1)/2$ ” (see page 2 of the amendment), it still lacks specific definition or description for N and  $N$ , so that it is unclear what the difference between N and  $N$  is and which of referenced letters is really used in the context. For example, “N” in lines 4-6 on page 4 is really means “N”, or “ $N$ ”? Appropriate correction or explanation is required.
  - c. On page 5, line 15, the term “ANWE” lacks an antecedent definition or description in the specification and is not commonly use term in the art, even though the applicant argues that “applicant is simply using an acronym for his invention titled ‘Automated New Word Extraction’” (the amendment: page 8, paragraph 8). Appropriate correction is required. The examiner suggests inserting the term “ANWE” after the terms of “Automated New Word Extraction” that first appear in the body of the specification, such as in line 10 of page 2,

### *Claim Rejections - 35 USC § 103*

5. Claims 1-3, 6-12 and 15-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wang et al. (US 6,904,402 B1) hereinafter referenced as Wang, in view of Razin et al. (US 6,098,034) hereinafter referenced as Razin.

As per **claim 1**, Wang discloses system and iterative method for lexicon, segmentation and language model joint optimization (title), comprising:

“segmenting a cleaned corpus to form a segmented corpus”, (Fig. 5 and column 9, lines 36-44, ‘segmentation’, ‘the received corpus is built’, ‘pre-processed to remove some obvious illogical words (so as to provide cleaned corpus)’);

“splitting the segmented corpus to form sub strings, and counting the occurrences of each sub strings appearing in the corpus” (column 1, lines 45-60, ‘a textual corpus is dissected (interpreted as split) into a plurality of items (sub strings)’ and ‘counts the number of occurrences of a particular item (word, character, etc.)’); and

Even though Wang further suggests that ‘the items of the corpus’ having low occurrence frequency ‘may be pruned’ (column 7, lines 27-29) and ‘counting the occurrence of strings of characters’ (corresponding to new words and is capable of outputting), Wang does not expressly disclose “**filtering** out false candidates to output new words”. However, this feature is well known in the art as evidenced by Razin who, in the same field of endeavor, discloses method for standardizing phrasing in a document (title), comprising ‘filtering the preliminary list of extracted phrases (candidates) to create (output) a final list of extracted phrases (corresponding new words)’ (Fig. 2 and column 29, lines 55-56). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Wang by specifically providing filtering a set of extracted phrases and creating (output) final phrases list, as taught by

Art Unit: 2654

Razin, for the purpose (motivation) of obtaining extracted words constituting significant user phrases (or new words) (Razin: column 2, lines 46-47).

As per **claim 2** (depending on claim 1), Wang in view of Razin further discloses “using punctuations, Arabic digits and alphabetic strings, or new words patterns to split the cleaned corpus”, (Razin, column 21, lines 10, ‘punctuation’; column 4, lines 26, ‘the usage of stop list’);

As per **claim 3** (depending on claim 1), Wang in view of Razin further discloses “using common vocabulary to segment the cleaned corpus”, (Razin: column 5, lines 36-45, ‘the dictionary of standard phrases (common vocabulary)’).

As per **claim 6** (depending on claim 1), Wang in view of Razin further discloses:

“filtering out functional words” (Razin: column 4, lines 35-38, ‘stop list’, ‘semantically insignificant words (e.g., “and then about the”) (interpreted as functional words)’, which suggests that these words can be filtered out);

“filtering out those sub strings which almost always appear along with a longer sub strings” (Razin: column 9, lines 52, ‘eliminates from the phrase list otherwise-significant phrases that are nested within other significant phrases... removes from the final phrase list minimal content words dangling at the beginning or end of preliminary user-specific phrases’, which reads on the claim); and

”filtering out those sub strings for which the occurrence is less than a predetermined threshold”, (Razin: column 2, lines 10-13, ‘each node of tree is associated with a record of the number of occurrence of the word sequence at that node, where the number of occurrence exceeds the required threshold’, which reads on the claimed limitation).



As per **claim 7** (depending on claim 1), Wang in view of Razin further discloses “using pre-recognized functional words as segment boundary patterns”, (Razin: column 4, lines 35-38, ‘stop list’, ‘semantically insignificant words (e.g., “and then about the”) (interpreted as functional words)’).

As per **claim 8** (depending on claim 3), the rejection is based on the same reason described for claim 7 because the claim recites the same or similar limitation(s) as claim 7.

As per **claim 9** (depending on claim 3), the rejection is based on the same reason described for claim 6 because the claim recites the same or similar limitation(s) as claim 6.

As per **claims 10-12 and 15-18**, they recite an automatic new word extraction system. The rejection is based on the same reason described for claims 1-3 and 6-9, respectively, because the claims recite the same or similar limitation(s) as claims 1-3 and 6-9, respectively.

As per **claim 19**, it recites a program storage device readable by machine. The rejection is based on the same reason described for claim 1, because the claim recites the same or similar limitations as claim 1.

6. Claims 4-5 and 13-14 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wang in view of Razin as applied to claims 1 and 10, and further in view of Hui (IDS: "Color Set Size Problem with Applications to String Matching," Proc. of 2nd Symposium on Combinatorial Pattern Matching, 1992, pp. 230-243).

As per **claim 4** (depending on claim 1), even Wang in view of Razin further discloses using suffix tree (i.e. atomic suffix tree—AST) (Wang: column 1, line 42; Razin: column, 2, line 3), Wang in view of Razin does not expressly disclose “using a GAST”. However, the feature is

Art Unit: 2654

well known in the art as evidenced by Hui who teaches ‘the concept of suffix tree can be extended’ and ‘this extension is called the Generalized suffix tree (GST)( corresponding to GAST)’ (Hui, page 237, first paragraph). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Wang in view of Razin by specifically providing using extended suffix tree (GST or GAST), for the purpose of storing more than one input strings (Hui: page 237, first paragraph).

As per **claim 5** (depending on claim 4), Wang in view of Razin and Hui further discloses the tree “implemented by limiting length of sub strings”, (Razin: column 14, lines 34-35, ‘length less than or equal to Smax’).

As per **claim 13** (depending on claim 10), the rejection is based on the same reason described for claim 4 because the claim recites the same or similar limitation(s) as claim 4.

As per **claim 14** (depending on claim 10), the rejection is based on the same reason described for claim 5 because the claim recites the same or similar limitation(s) as claim 5.

### ***Conclusion***

7. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR

Art Unit: 2654

1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

8. Please address mail to be delivered by the United States Postal Service (USPS) as follows:

Mail Stop \_\_\_\_\_  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**or faxed to:** 571-273-8300, (for formal communications intended for entry)

**Or:** 571-273-8300, (for informal or draft communications, and please label "PROPOSED" or "DRAFT")

If no Mail Stop is indicated below, the line beginning Mail Stop should be omitted from the address.

Effective January 14, 2005, except correspondence for Maintenance Fee payments, Deposit Account Replenishments (see 1.25(c)(4)), and Licensing and Review (see 37 CFR 5.1(c) and 5.2(c)), please address correspondence to be delivered by other delivery services (Federal Express (Fed Ex), UPS, DHL, Laser, Action, Purolater, etc.) as follows:

U.S. Patent and Trademark Office  
Customer Window, Mail Stop \_\_\_\_\_  
Randolph Building  
Alexandria, VA 22314

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Qi Han whose telephone numbers is (571) 272-7604. The examiner can normally be reached on Monday through Thursday from 9:00 a.m. to 7:00 p.m. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil, can be reached on (571) 272-7602.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Inquiries regarding the status of submissions relating to an application or questions on the Private PAIR system should be directed to the Electronic Business Center (EBC) at 866-217-9197 (toll-free) or 703-305-3028 between the hours of 6 a.m. and midnight Monday through Friday EST, or by e-mail at: [ebc@uspto.gov](mailto:ebc@uspto.gov). For general information about the PAIR system, see <http://pair-direct.uspto.gov>.

QH/qh  
July 25, 2005

  
RICHEMOND DORVIL  
SUPERVISORY PATENT EXAMINER